

Utilisation des bases de données Espadon pour projet de visualisation

Nous utilisons les données du logiciel Espadon de la FFN pour permettre une visualisation facile et ergonomique de ces données. La visualisation est accessible sur le notebook suivant : <https://observablehq.com/d/4ab4aeb32d243713>

Notre notebook lit les données au format CSV. Or, les données d'Espadon sont enregistrées au format *Microsoft Access*. De plus, elles suivent un schéma relationnel quelque peu complexe. Facilitant l'exploitation logicielle mais pas la visualisation de données. Enfin, chaque compétition a sa propre base de données.

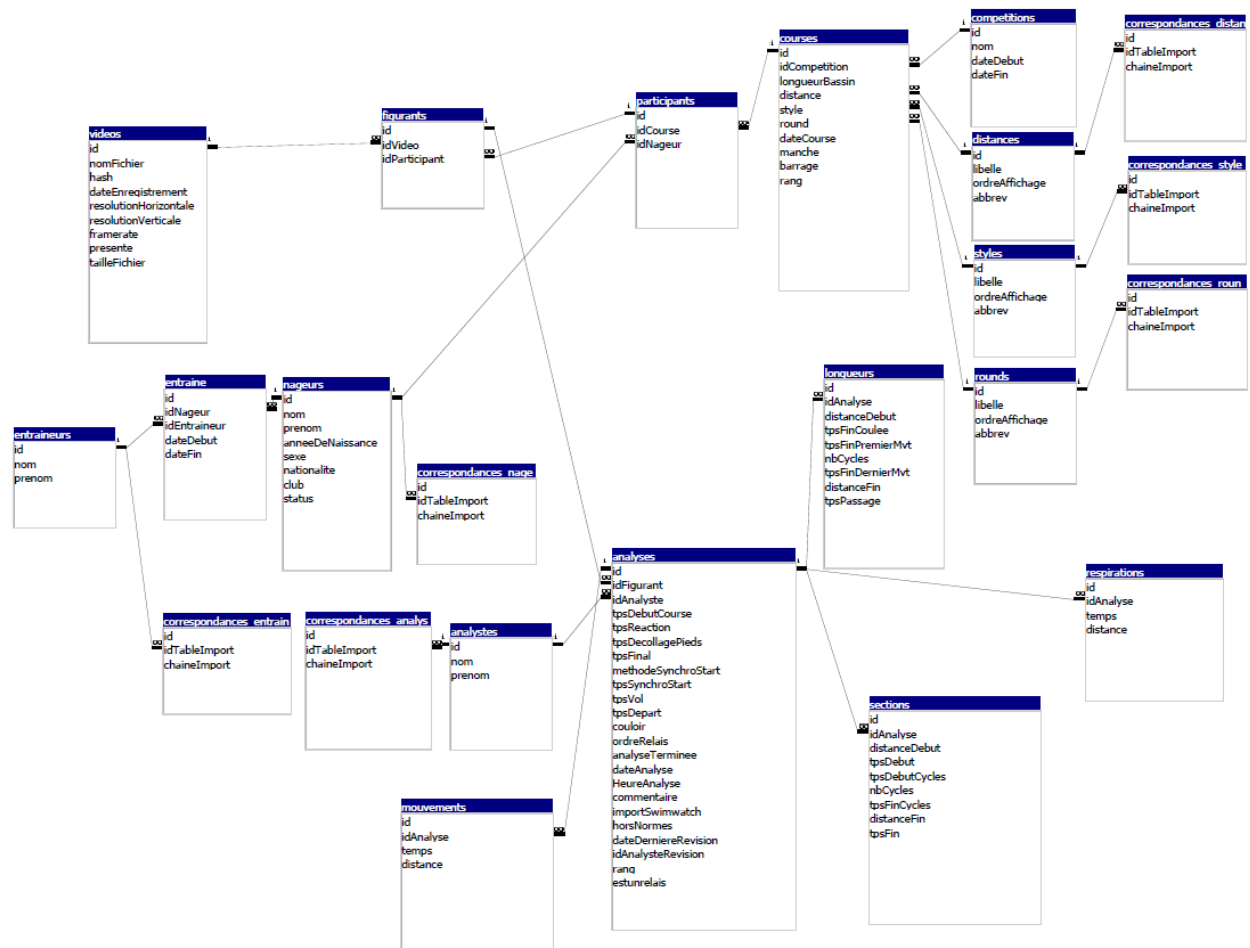
Pour utiliser ces données facilement nous devons :

1. Réorganiser les données selon un nouveau schéma
2. Modifier le format
3. Regrouper toutes les données au même endroit

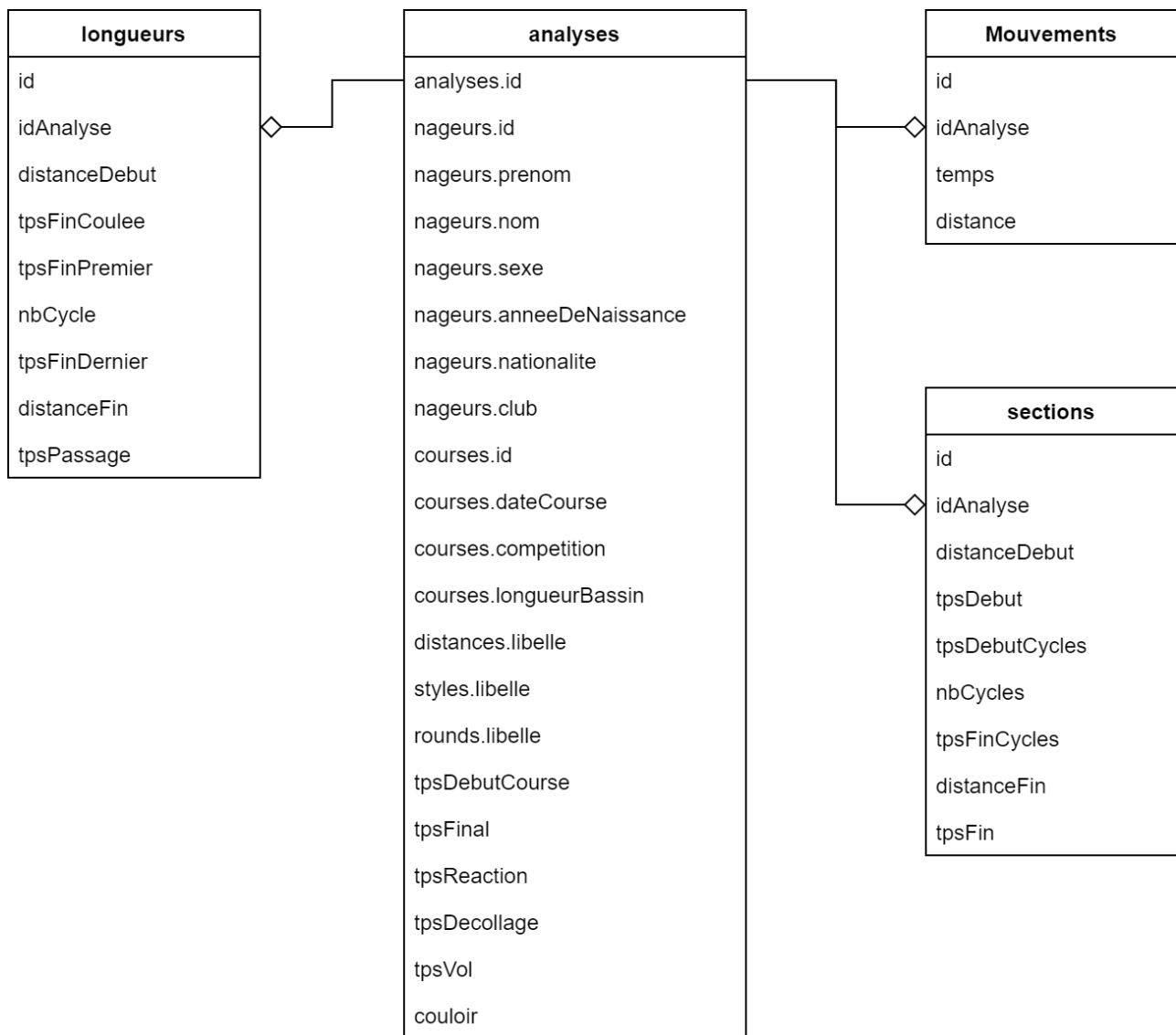
Réorganisation des données

temps estimé : ~ 2 min par BDD

Les bases de données issues du logiciel Espadon suivent le schéma relationnel suivant :



Nous préférons manipuler une base de données avec moins de relation. Une jointure multiple des tables *analyses*, *figurants*, *participants*, *nageurs*, *courses*, *distances*, *styles* et *rounds* donne les relations suivantes :



Ce schéma promet une exploitation plus facile. Pour obtenir la nouvelle table *analyse*, il s'agit d'exécuter la commande SQL suivante dans *Microsoft Access* :

```
SELECT analyses.id, nageurs.id, nageurs.prenom, nageurs.nom,
nageurs.sexe, nageurs.anneeDeNaissance, nageurs.nationalite,
nageurs.club, courses.id, courses.dateCourse, competitions.nom,
courses.longueurBassin, distances.libelle, styles.libelle,
rounds.libelle, analyses.tpsDebutCourse, analyses.tpsFinal,
analyses.tpsReaction, analyses.tpsDecollagePieds, analyses.tpsVol,
analyses.couloir
FROM styles
INNER JOIN rounds
    INNER JOIN ((nageurs
        INNER JOIN (distances
            INNER JOIN ((competitions
                INNER JOIN courses
                    ON competitions.id =
courses.idCompetition))
```

```

INNER JOIN participants
ON courses.id =
participants.idCourse)
ON distances.id = courses.distance)
ON nageurs.id = participants.idNageur)
INNER JOIN (figurants
INNER JOIN analyses
ON figurants.id = analyses.idFigurant)
ON participants.id = figurants.idParticipant)
ON rounds.id = courses.round)
ON styles.id = courses.style;

```

Les 3 autres tables ne sont pas modifiées.

Modification du format

temps estimé : ~ 4 min par BDD

Pour convertir les données en CSV, pour les quatre tables qui nous intéressent :

- Copier la table dans *Access*.
- Coller la table dans *Excel*.
- Remplacer les virgules des nombres réels par des points. **CTRL + H** pour rechercher et tout remplacer.
- Enregistrer au format UTF-8 CSV

Pour faciliter la suite du processus, enregistrer les fichiers avec les noms : *analyses.csv*, *longueurs.csv*, *mouvements.csv*, *sections.csv*.

Fusions des différentes bases de données

temps estimé : ~ 30 s par ajout de BDD

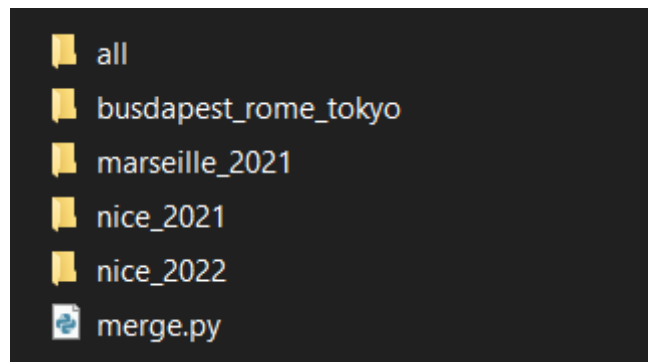
L'idée pour fusionner les bases de données est de mettre bout à bout les fichiers CSV extraits des bases de données. Le problème qui s'expose est le suivant :

Les BDD peuvent avoir des mêmes identifiants d'analyses, de courses et du nageurs sans pour autant désigner les mêmes ressources. Il faut donc assurer l'unicité des identifiants et conserver tout en gardant le lien entre les quatre tables. Ou au contraire désigner une même ressource avec des identifiants différents (les nageurs par exemple).

Pour y répondre, nous avons développé un script Python *merge.py* . A chaque ajout d'une table, il permet de modifier les identifiants s'il sont déjà utilisé. Il assure que *idAnalyse* soit modifié dans les autres tables également. Et réutilise l'identifiant d'un nageur si celui-ci apparaît déjà.

Pour l'utiliser, se placer dans le même répertoire que le script :

- Avoir un repertoire *all* qui contiendra les fusions des fichiers
- Dans un répertoire au nom de la BDD, mettre 4 fichiers CSV
- Exécuter le script. Lorsqu'il demande *nom de la BDD ?* indiquer le nom de répertoire de la BDD à ajouter.
- Récupérer les fichiers dans *all*



```
PS D:\Github\PAR_SwimAnalysis\DataObservable> python merge.py  
Nom de la BDD à ajouter : marseille_2021
```